Foto 1
Curso 1988-1989

Estadísticos de tendencia central y de dispersión. Media y varianza

(1) Introducción

La media aritmética o promedio ($\bar{X}$) y la varianza (S^2), ó su raíz cuadrada la desviación típica (S) son los estadísticos fundamentales en el análisis estadístico en Biología.

Supongamos que hemos medido una magnitud en una muestra de 5 individuos (por ej: la variable que estudiamos es la longitud de la corola en centrímetros).

Las distintas modalidades o los distintos valores que toma ese caracter o variable en los 5 individuos son por ejemplo.

15, 16, 18, 20, 21 cm

El valor del caracter X (longitud de corola) en cada individuo se denomina con subíndices

X_1, X_2, X_3, X_4, X_5

En general X_i son los valores que toma la variable X en cada individuo

¿Cómo tendremos una visión de conjunto de la información contenida en esa muestra?

En primer lugar, tenemos que advertir que la muestra es pequeña ($n=5$) y que cada valor sólo aparece 1 vez (frecuencia -fi- de cada valor $= 1$)

Por tanto para resumir esa información.

- no necesitamos resumirlo en tabla de distribución de frecuencias (n pequeño y fi de cada valor $= 1$)
- no necesitamos hacer representaciones gráficas (por la misma razón).
- sólo tendremos que "condensar" dicha información en algunos números que la expresen de forma clara y concisa *

La información contenida en una muestra queda perfectamente condensada o sintetizada por el cálculo de unos índices que la definen. Estos índices se denominan estadísticos (por ser muestrales) y pueden ser de centralización y de dispersión siendo la media aritmética y la varianza respectivamente los representantes más utilizados.

> "Así pues con la media y la varianza queda definida la información contenida en una muestra"

Fotocopia nº 5
curso 1988-1989

Ejemplo nº 1 Cálculo de $\overline{X}$, y propiedades de la media

A) Se ha medido la longitud del cuerpo (en milímetros) en 5 individuos de una determinada especie de ácaro

X_i (long. cuerpo) → 15, 16, 18, 20, 21.

$\overline{X} = \frac{\sum_{i=1} X_i}{n} = \frac{90}{5} = 18$ mm. (← Fórmula de $\overline{X}$ para datos no agrupados por frecuencias)

a) ¿qué ocurre si a todos los datos de esta muestra sumamos una constante, por ejemplo 10? (1ª propiedad)

(X_i+10) → 25, 26, 28 30, 31

$\overline{X}_{(X_i+10)} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{140}{5} = 28$ ← Observad que 28 = 18 + 10 = $\overline{X}$ + constante

b) ¿qué ocurre si a todos los datos de esta muestra restamos una constante, por ejemplo 10? (2ª propiedad).

(X_i-10) → 5, 6, 8 10, 11

$\overline{X}_{(X_i-10)} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{40}{5} = 8$ ← Observad que 8 = 18 − 10 = $\overline{X}$ − constante

c) ¿qué ocurre si a todos los datos de esta muestra multiplicamos una constante, por ej 10? (3ª propiedad).

$(X_i \cdot 10)$ → 150, 160, 180, 200, 210

$\overline{X}_{(X_i \cdot 10)} = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{900}{5} = 180$ ← observad que 180 = 18 · 10 = $\overline{X}$ · constante

d) ¿qué ocurre si a todos los datos de esta muestra dividimos una constante, por ej. 10? (4ª propiedad)

$(X_i/10)$ → 1'5, 1'6, 1'8, 2, 2'1

$\overline{X}_{(X_i/10)} = \frac{\sum X_i}{n} = \frac{9}{5} = 1'8$ — Observad que 1'8 = 18/10 = $\overline{X}$/constante.

B) Se ha medido la longitud del cuerpo (en milímetros) en 14 individuos de una determinada especie de ácaro.

X_i	f_{abs}	$X_i f_{ab}$
15	2	30
16	3	48
18	4	72
20	3	60
21	2	42
	N=14	252

$\overline{X} = \frac{\sum_{i=1}^{n} X_i f_i}{n} = \frac{252}{14} = 18$ mm (← Fórmula de $\overline{X}$ para datos agrupados por frecuencias).

Compruebe en este caso, las propiedades de la media.
(X_i + constante)
(X_i − constante)
(X_i · constante)
(X_i / constante)

1. Cálculo de media aritmetica ($\overline{X}$)

1-a) Para datos no agrupados por frecuencias

$$\overline{X} = \frac{1}{n}(x_1 + x_2 + \dots x_n) = \frac{1}{n}\sum_{i=1}^{n} x_i = \boxed{\frac{\sum_{i=1}^{n} x_i}{n}}$$

1-b) Para datos agrupados por frecuencias.

$$\overline{X} = \frac{1}{n}(x_1 f_1 + x_2 f_2 + \dots x_n f_n) = \frac{1}{n}\sum_{i=1}^{n} x_i f_i = \boxed{\frac{\sum_{i=1}^{n} x_i f_i}{n}}$$

2. Propiedades de la media ($\overline{x}$) (codificación de datos)

2-a) Cuando a todos los datos (x_i) de una muestra se les suma una constante (a), la media ($\overline{X}$) queda sumada a esa constante (a).

$$\overline{X}(x_i+a) = \frac{(x_1+a)+(x_2+a)+\dots+(x_n+a)}{n} = \frac{\sum_{i=1}^{n}(x_i+a)}{n} = \frac{\sum_{i=1}^{n} x_i + na}{n} = \frac{\sum x_i}{n} + \frac{na}{n} = \overline{x} + a$$

2-b) Cuando a todos los datos (x_i) de una muestra se les resta una constante (a), la media ($\overline{X}$) queda restada a esa constante (a)

$$\overline{X}(x_i-a) = \frac{(x_1-a)+(x_2-a)+\dots(x_n-a)}{n} = \frac{\sum_{i=1}^{n}(x_i-a)}{n} = \frac{\sum_{i=1}^{n} x_i - na}{n} = \frac{\sum_{i=1}^{n} x_i}{n} - \frac{na}{n} = \overline{X} - a$$

2-c) Cuando a todos los datos (x_i) de una muestra se les multiplica por una constante (a), la media ($\overline{X}$) queda multiplicada por esa constante (a).

$$\overline{X}(x_i \cdot a) = \frac{(x_1\, a)+(x_2\, a)+\dots(x_n \cdot a)}{n} = \frac{\sum_{i=1}^{n}(x_i\, a)}{n} = \frac{a\sum_{i=1}^{n} x_i}{n} = a \cdot \overline{X}$$

2-d) Cuando a todos los datos (x_i) de una muestra se les divide por una constante (a), la media ($\overline{x}$) queda dividida por esa constante (a)

$$\overline{X}(x_i/a) = \frac{x_1/a + x_2/a + \dots x_n/a}{n} = \frac{\sum_{i=1}^{n} x_i/a}{n} = \frac{1/a \sum_{i=1}^{n} x_i}{n} = \frac{\sum_{i=1}^{n} x_i}{a \cdot n} = \frac{\overline{X}}{a}$$

---

1. Cálculo de la varianza (s^2)

1-a) Para datos no agrupados por frecuencias.

$$S^2 = \boxed{\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1}} = \frac{\sum_{i=1}^{n}(x_i^2 - 2x_i\overline{X} + \overline{X}^2)}{n-1} \overset{(*)}{=} \frac{\sum_{i=1}^{n} x_i^2 - 2\overline{x}\sum_{i=1}^{n} x_i + n\overline{X}^2}{n-1} = \frac{\sum_{i=1}^{n} x_i^2 - 2\frac{(\sum x_i)^2}{n} + n\frac{(\sum x_i)^2}{n^2}}{n-1} =$$

$$= \boxed{\frac{\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2/n}{n-1}} = \frac{SC}{gl}$$

1-b) Para datos agrupados por frecuencias.

$$S^2 = \boxed{\frac{\sum_{i=1}^{n} f_i(x_i-\overline{X})^2}{n-1}} = \frac{\sum_{i=1}^{n} f_i x_i^2 - 2\overline{X}\frac{(\sum_{i=1}^{n} x_i f_i)^2}{n} + \frac{(\sum_{i=1}^{n} x_i f_i)^2}{n}}{n-1} = \boxed{\frac{\sum_{i=1}^{n} x_i^2 f_i - (\sum_{i=1}^{n} x_i f_i)^2/n}{n-1}}$$

2. Propiedades de la varianza (s^2) (codificación de datos)

2-a) Cuando a todos los datos (x_i) de una muestra se les suma una constante (a), la varianza (S^2) no varía.

$$S^2(x_i+a) = \frac{\sum_{i=1}^{n}\left[(x_i+a) - \left[\sum_{i=1}^{n}(x_i+a)/n\right]\right]^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i + a - \sum_{i=1}^{n} x_i/n - na/n)^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i + a - \overline{X} - a)^2}{n-1} = \frac{\sum(x_i-\overline{x})^2}{n-1} = S^2.$$

2-b) Cuando a todos los datos (x_i) de una muestra se les resta una constante (a), la varianza (S^2) no varía.

$$S^2(x_i-a) = \frac{\sum_{i=1}^{n}\left[(x_i-a) - \left[\sum_{i=1}^{n}(x_i-a)/n\right]\right]^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i - a - \sum_{i=1}^{n} x_i/n + na/n)^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i - a - \overline{x} + a)^2}{n-1} = \frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1} = S^2$$

2-c) Cuando a todos los datos (x_i) de una muestra se les multiplica por una constante (a), la varianza (S^2) queda multiplicada por esa constante al cuadrado

$$S^2(x_i \cdot a) = \frac{\sum_{i=1}^{n}\left[a.x_i - (\sum_{i=1}^{n} a x_i/n)\right]^2}{n-1} = \frac{\sum_{i=1}^{n}\left[a x_i - a\sum_{i=1}^{n} x_i/n\right]^2}{n-1} = \frac{\sum_{i=1}^{n}\left[a(x_i-\overline{x})\right]^2}{n-1} = \frac{\sum_{i=1}^{n} a^2(x_i-\overline{x})^2}{n-1} = \frac{a^2\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1} = a^2 S^2$$

2-d) Cuando a todos los datos (x_i) de una muestra se les divide por una constante (a), la varianza (S^2) queda dividida por esa constante al cuadrado.

$$S^2(x_i/a) = \frac{\sum_{i=1}^{n}\left[x_i/a - \sum_{i=1}^{n} x_i/a/n\right]^2}{n-1} = \frac{\sum_{i=1}^{n}\left(\frac{1}{a}x_i - \frac{1}{a}\sum_{i=1}^{n} x_i/n\right)^2}{n-1} = \frac{\sum_{i=1}^{n}\left(\frac{1}{a}x_i - \frac{1}{a}\overline{X}\right)^2}{n-1} = \frac{\sum_{i=1}^{n}\left[\frac{1}{a}(x_i-\overline{x})\right]^2}{n-1} = \frac{\sum_{i=1}^{n}\frac{1}{a^2}(x_i-\overline{x})^2}{n-1} = \frac{\frac{1}{a^2}\sum_{i=1}^{n}(x_i-\overline{x})^2}{n-1} = \frac{S^2}{a^2}$$

Fotocopia (nº 6) curso 198?-1989

Ejemplo nº 2 Calculo de S^2 y S, propiedades de la varianza. (para datos no agrupados por frecuencias)

A) Se ha medido la longitud del cuerpo (en milímetros) en 5 individuos de una determinada especie de ácaro.

X_i (long. cuerpo) → 15, 16, 18, 20, 21. $\bar{X} = \frac{90}{5} = 18$

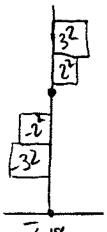
$$S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{N-1} \quad \text{--- transformación (incluida en Fot nº 4 ✱) ---→} \quad \frac{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2/n}{n-1}$$

1- procedimiento aritmético.

X_i	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
15	−3	9
16	−2	4
18	0	0
20	2	4
21	3	9
		26

$$S^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{N-1} = \frac{SC}{gl} = \frac{26}{4} = 6'5$$

2- procedimiento geométrico.



$$S^2 = \frac{\square + \square + \square + \square}{4} = \frac{SC}{gl} = \frac{26}{4} = 6'5$$

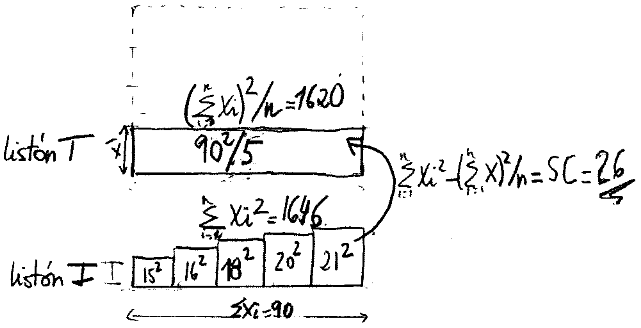
$$S = \sqrt{S^2} = \sqrt{6'5} = 2'55$$

1- procedimiento aritmético

X_i	X_i^2
15	225
16	256
18	324
20	400
21	441
90	1646

$$S^2 = \frac{\sum_{i=1}^{n} X_i^2 - (\sum_{i=1}^{n} X_i)^2/n}{N-1} = \frac{1646 - 1620}{4} = \frac{26}{4} = 6'5$$

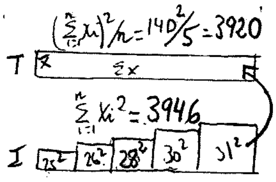
2. procedimiento geométrico.



$$S^2 = \frac{SC}{gl} = \frac{\text{listón I} - \text{listón T}}{gl} = \frac{26}{4} = 6'5$$

a) ¿qué ocurre si a todos los individuos de la muestra sumamos una constante, por ej. 10? (propiedad 1).

$(X_i + 10)$ → 25, 26, 28, 30, 31



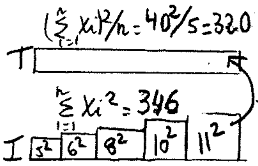
I−T = SC = 26.

$$S^2_{(X_i+10)} = \frac{SC}{gl} = \frac{26}{4} = 6'5$$

observad que la S^2 no varía

b) ¿qué ocurre si a todos los individuos de la muestra restamos una constante, por ej. 10? (propiedad 2).

$(X_i - 10)$ → 5, 6, 8, 10, 11



I−T = SC = 26

$$S^2_{(X_i-10)} = \frac{SC}{gl} = \frac{26}{4} = 6'5$$

observad que la S^2 no varía

c) ¿qué ocurre si a todos los individuos de la muestra multiplicamos por una constante, por ej 10 (propiedad 3)

$(X_i \cdot 10) \rightarrow 150, 160, 180, 200, 210$

T: $\left(\sum_{i=1}^{n} X_i\right)^2/N = 900^2/5 = 162.000$

I: $\sum_{i=1}^{n} X_i^2 = 164.600$ (150^2, 160^2, 180^2, 200^2, 210^2)

$I - T = 2600$

$S^2_{(X_i \cdot 10)} = \frac{2600}{4} = 650$

observad que $650 = 6'5 \cdot 10^2$
$= S^2 \cdot \text{constante}^2$

d) ¿que ocurre si a todos los individuos de la muestra dividimos por una constante, por ej 10 (propiedad 4)

$(X_i/10) \rightarrow 1'5, 1'6, 1'8, 2, 2'1$

$\left(\sum_{i=1}^{n} X_i\right)^2/N = 9^2/5 = 16'2$

$\sum_{i=1}^{n} X_i^2 = 16'46$ ($1'5^2$, $1'6^2$, $1'8^2$, 2^2, $2'1^2$)

$I - T = 0'26$

$S^2_{(X_i/10)} = \frac{0'26}{4} = 0'065$

observad que $0'065 = \frac{6'5}{10^2}$
$= S^2/\text{constante}^2$

Ejemplo:

Calcular la $\bar{X}$, S^2 y comprobar propiedades de ambas en la siguiente muestra que representa la anchura de la corola (en mm) medida en 7 plantas de la misma especie

$X_i \rightarrow 4, 6, 8, 10, 12, 14, 16$

Fotocopia nº 8
curso 1988-1989

Ejemplo nº 3. Cálculo de S^2 y S, propiedades de la varianza (para datos agrupados por frecuencias)

B) Se ha medido la longitud del cuerpo (en milímetros) en 14 individuos de una determinada especie de ácaro.

x_i	f_i
15	2
16	3
18	4
20	3
21	2

$$\bar{X} = \frac{\sum_{i=1}^{14} x_i f_i}{N} = \frac{252}{14} = 18 \text{ mm}$$

Calculemos la varianza (S^2) por los procedimientos que conocemos

fórmula conceptual

$$S^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{N-1}$$

= – – transformación incluida en Fot nº – – > =

fórmula transformada para el cálculo

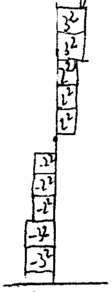
$$\frac{\sum_{i=1}^{n} X_i^2 f_i - \left(\sum_{i=1}^{n} X_i f_i\right)^2 / N}{N-1}$$

1 – Procedimiento aritmético

x_i	fa.	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$	$fa(x_i - \bar{x})^2$
15	2	−3	9	18
16	3	−2	4	12
18	4	0	0	0
20	3	2	4	12
21	2	3	9	18
				60

$$S^2 = \frac{\sum_{i=1}^{n} f_i (x_i - \bar{x})^2}{N-1} = \frac{60}{13} = 4'6$$

2 – Procedimiento geométrico



$\bar{x} = 18$

$$S^2 = \frac{\square + \square + \square + \diamond + \diamond + \diamond + \diamond + \diamond + \diamond + \diamond}{N-1} = \frac{SC}{gl} = \frac{60}{13} = 4'6$$

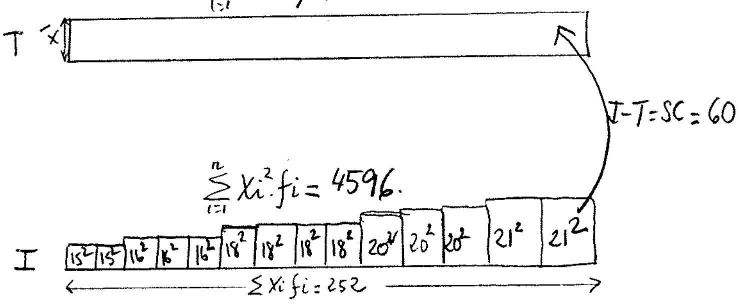
1 – Procedimiento aritmético

x_i	fa	x_i^2	$x_i^2 fa$	$x_i fa$
15	2	225	450	30
16	3	256	768	48
18	4	324	1296	72
20	3	400	1200	60
21	2	441	882	42
90	14		4596	252

$$S^2 = \frac{\sum_{i=1}^{n} x_i^2 f_i - \left(\sum_{i=1}^{n} x_i f_i\right)^2 / N}{N-1} = \frac{4596 - 252^2/14}{13} = \frac{60}{13} = 4'6$$

2 – Procedimiento geométrico

$\sum_{i=1}^{n} (x_i f_i)^2 / 14 = 252^2/14 = 4536$



$\sum_{i=1}^{n} X_i^2 f_i = 4596$

$$S^2 = \frac{SC}{gl} = \frac{\text{listón I} - \text{listón T}}{gl} = \frac{60}{13} = 4'6$$

Compruebe en este caso, las propiedades de la varianza
(x_i + constante)
(x_i − constante)
(x_i · constante)
(x_i / constante)

Fotocopia nº 9
curso 1988-1989

Ejemplo nº 4: Estudio de la misma variable en 3 muestras de distinto tamaño

Se ha medido la longitud del cuerpo en una determinada especie de ácaro en 3 muestras

Muestra 1 (n = 11)

x_i	f_i	$x_i f_i$
15	1	15
16	2	32
19	4	76
20	3	60
21	1	21
	11	204

$\bar{X}_1 = \frac{\sum_{i=1}^{11} x_i f_i}{n} = \frac{204}{11} = 18'5$

Muestra 2 (n = 16)

x_i	f_i	$x_i f_i$
15	1	15
16	2	32
17	3	51
19	4	76
20	3	60
22	2	44
23	1	23
	16	301

$\bar{X}_2 = \frac{\sum_{i=1}^{16} x_i f_i}{16} = \frac{301}{16} = 18'8$

Muestra 3 (n = 6)

x_i	f_i	$x_i f_i$
15	1	15
17	2	34
18	2	36
19	1	19
	6	104

$\bar{X}_3 = \frac{\sum_{i=1}^{6} x_i f_i}{6} = \frac{104}{6} = 17'3$

(1) $$\bar{X}_{\bar{x}_1+\bar{x}_2+\bar{x}_3 \text{ ponderada}} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3}{n_1+n_2+n_3} = \frac{11 \cdot 18'5 + 16 \cdot 18'8 + 6 \cdot 17'3}{11+16+6} =$$

$$= \frac{203'5 + 300'8 + 103'8}{33} = \frac{608'1}{33} = 18'4$$

(Si $n_1 = n_2 = n_3$)

$$\bar{X}_{\bar{x}_1+\bar{x}_2+\bar{x}_3} = \frac{n\bar{X}_1 + n\bar{X}_2 + n\bar{X}_3}{n+n+n} = \frac{n(\bar{X}_1+\bar{X}_2+\bar{X}_3)}{3n} = \frac{\bar{X}_1+\bar{X}_2+\bar{X}_3}{3}$$

(2) $$S^2_{s^2_1+s^2_2+s^2_3 \text{ ponderada}} = \frac{gl_1 \cdot S^2_1 + gl_2 \cdot S^2_2 + gl_3 \cdot S^2_3}{gl_1+gl_2+gl_3} = \frac{SC_1 + SC_2 + SC_3}{a(n-1)}$$

a = nº de muestras
n = tamaño de muestra

→ "Calculad la varianza ponderada de las 3 muestras"

(3) $$S_{s_1+s_2+s_3 \text{ ponderada}} = \sqrt{S^2 \text{ponderada}}$$

Ejemplo nº 5: Estudio de distintas variables en 3 muestras. Coeficiente de variación (CV)

Muestra 1

variable (long. cuerpo)

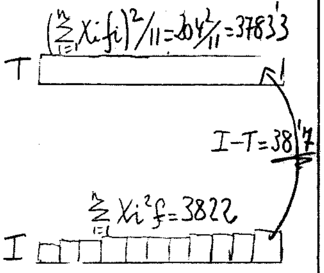
x_i	f_i	$x_i f_i$
15	1	15
16	2	32
19	4	76
20	3	60
21	1	21
	11	204

$\bar{X}_1 = \dfrac{\sum_{i=1}^{n} x_i f_i}{n} = \dfrac{204}{11} =$ 18'5

$S_1^2 = \dfrac{\sum_{i=1}^{n} x_i^2 f_i - \left(\sum_{i=1}^{n} x_i f_i\right)^2 / n}{n-1} =$

$= \dfrac{\text{listón I} - \text{listón T}}{n-1} =$

= 38'7 / 10 = 3'9



$S_1 = \sqrt{S_1^2} = \sqrt{3'9} =$ 1'98

$CV_1 = \dfrac{S_1}{\bar{X}_1} =$ 1'98 / 18'5 = 0'11

expresado en % → 11%

Muestra 2

variable (long. ala)

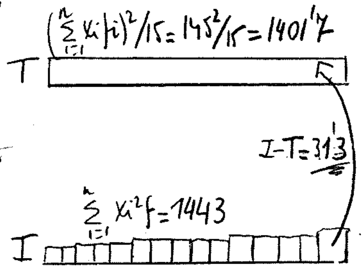
x_i	f_i	$x_i f_i$
7	2	14
8	3	24
10	5	50
11	4	44
13	1	13
	15	145

$\bar{X}_2 = \dfrac{\sum_{i=1}^{n} x_i f_i}{n} = \dfrac{145}{15} =$ 9'7

$S_2^2 = \dfrac{\sum_{i=1}^{n} x_i^2 f_i - \left(\sum_{i=1}^{n} x_i f_i\right)^2 / n}{n-1} =$

$= \dfrac{\text{listón I} - \text{listón T}}{n-1} =$

= 31'3 / 14 = 2'2



$S_2 = \sqrt{S_2^2} = \sqrt{2'2} =$ 1'48

$CV_2 = \dfrac{S_2}{\bar{X}_2} =$ 1'48 / 9'7 = 0'15

expresado en % → 15%

Muestra 3

variable (long. pata)

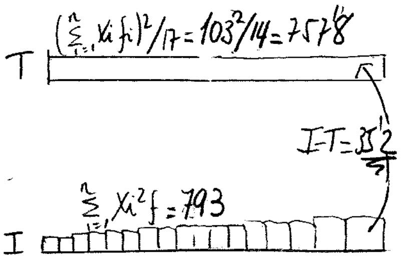
x_i	f_i	$x_i f_i$
5	2	10
6	4	24
8	4	32
9	3	27
10	1	10
	14	103

$\bar{X}_3 = \dfrac{\sum_{i=1}^{n} x_i f_i}{n} =$ 7'4

$S_3^2 = \dfrac{\sum_{i=1}^{n} x_i^2 f_i - \left(\sum_{i=1}^{n} x_i f_i\right)^2 / n}{n-1} =$

$= \dfrac{\text{listón I} - \text{listón T}}{n-1} =$

= 35'2 / 14 = 2'5



$S_3 = \sqrt{S_3^2} = \sqrt{2'5} =$ 1'58

$CV_3 = \dfrac{S_3}{\bar{X}_3} =$ 1'58 / 7'4 = 0'21

expresado en % → 21%